

**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ**  
Федеральное государственное бюджетное  
образовательное учреждение высшего образования  
«**ТЮМЕНСКИЙ ИНДУСТРИАЛЬНЫЙ УНИВЕРСИТЕТ**»

**УТВЕРЖДАЮ**

Заведующий кафедрой ИСТ

\_\_\_\_\_ О. Ф. Данилов

« \_\_\_\_\_ » \_\_\_\_\_ 20\_\_ г.

**ФОНД ОЦЕНОЧНЫХ СРЕДСТВ**

дисциплины: **Модели и методы интеллектуального анализа**

направление подготовки: **09.04.04 Программная инженерия**

направленность (профиль): **Программная инженерия систем искусственного интеллекта**

форма обучения: **очная**

Фонд оценочных средств рассмотрен на заседании кафедры Интеллектуальных систем и технологий для направления 09.04.04 Программная инженерия направленность (профиль) Программная инженерия систем искусственного интеллекта

## 1. Формы аттестации по дисциплине

1.1. Форма промежуточной аттестации: экзамен.

Способ проведения промежуточной аттестации: устный опрос.

1.2. Формы текущей аттестации:

Таблица 1.1

№ п/п	Форма обучения	
	ОФО	
1.	Тестирование	
2.	Вопросы к защите отчетов по лабораторным работам	
3.	Защита домашней работы	
4.	Выполнение коллоквиума	

## 2. Результаты обучения по дисциплине, подлежащие проверке при проведении текущей и промежуточной аттестации

Таблица 2.1

№ п/п	Структурные элементы дисциплины		Код результата обучения по дисциплине	Оценочные средства	
	Номер раздела	Дидактические единицы (предметные темы)		Текущая аттестация	Промежуточная аттестация
1.	1	Введение, основные понятия анализа данных	31, 32, 33	Тесты	Вопросы к экзамену
2.	2	Математические объекты и методы в анализе данных	31, У1, В1	Вопросы к защите отчета по лабораторным работам	Вопросы к экзамену
3.	3	Линейная регрессия и классификация	31, У1, В1, 32, У2, В2, 33, У3, В3	Вопросы к защите отчета по лабораторным работам. Тесты. Домашнее задание	Вопросы к экзамену
4.	4	Логические методы	31, У1, В1, 32, У2, В2, 33, У3, В3	Вопросы к защите отчета по лабораторным работам	Вопросы к экзамену
5.	5	Особенности реальных данных	31, У1, В1, 32, У2, В2, 33, У3, В3	Вопросы к защите отчета по лабораторным работам. Коллоквиум.	Вопросы к экзамену
6.	6	Кластеризация данных	31, У1, В1, 32, У2, В2, 33, У3, В3	Вопросы к защите отчета по лабораторным работам. Домашнее задание. Тесты.	Вопросы к экзамену
7.	7	Нейронные сети	31, У1, В1, 32, У2, В2, 33, У3, В3	Вопросы к защите отчета по лабораторным работам. Тесты	Вопросы к экзамену

### 3. Фонд оценочных средств

3.1. Фонд оценочных средств, позволяющие оценить результаты обучения по дисциплине, включает в себя оценочные средства для текущей аттестации и промежуточной аттестации.

3.2. Фонд оценочных средств для текущей аттестации включает:

- комплект вопросов к защите отчетов по лабораторным работам - 75 шт., размещен в Приложении 1.

- примеры тестовых заданий - 56 шт. размещены в Приложении 2.

- комплект типовых заданий для выполнения домашней работы по теме "Линейная регрессия и классификация" - 1 шт., по теме "Кластеризация данных" - 1 шт. размещены в Приложении 3.

- комплект заданий к коллоквиуму по теме «Особенности реальных данных» - 10 шт. размещены в Приложении 4.

3.3. Фонд оценочных средств для промежуточной аттестации включает:

- комплект вопросов к экзамену по дисциплине Модели и методы интеллектуального анализа - 24 шт., размещен в Приложении 5.

**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ  
ФЕДЕРАЦИИ**

Федеральное государственное бюджетное  
образовательное учреждение высшего образования  
«ТЮМЕНСКИЙ ИНДУСТРИАЛЬНЫЙ УНИВЕРСИТЕТ»

**Комплект вопросов к защите отчетов по лабораторным работам**

по дисциплине "Модели и методы интеллектуального анализа"

**Раздел 2. Математические объекты и методы в анализе данных.**

**Лабораторная работа №1.** Анализ данных с помощью Pandas.

**Цель работы:** научиться пользоваться библиотекой Pandas и её встроенными объектами для визуализации данных в датасетах.

**Контрольные вопросы:**

1. Охарактеризуйте три вида методов интеллектуального анализа данных и приведите графическую интерпретацию их соотношения.
2. Какие методы образуют направление интеллектуального анализа данных?
3. Предложите дополнительные методы (свой вариант) для расширения совокупности методов направления интеллектуального анализа данных.
4. При решении каких проектных задач востребованы методы интеллектуального анализа данных?

**Раздел 3. Линейная регрессия и классификация.**

**Лабораторная работа №2.** Прогнозирование на основе статистического подхода.

**Цель работы:** изучить задачу прогнозирования временных рядов на примере применения статистических моделей.

**Контрольные вопросы:**

1. Постановка задачи, основные задачи анализа ВР. Критерии качества моделей. Стационарные и нестационарные временные ряды.
2. Какие основные классы методов анализа ВР? Data-driven и model-driven методы анализа. Проблемы прогнозирования.
3. Принципы прогнозирования в статистическом подходе к анализу ВР.
4. Декомпозиция ВР, типы паттернов.
5. Модели тренда ВР (на основе функций от времени).
6. Модели случайной компоненты ВР (AR, MA, ARMA, ARIMA).
7. Модели сезонных колебаний (индексные методы, адаптивные методы EST, спектральные методы, сезонная ARIMA).

**Лабораторная работа №3.** Классификация данных методом "К-ближайших соседей".

**Цель работы:** изучить метод простейший метод классификации данных «К-ближайших соседей» и научиться производить оценку данных с помощью визуальных инструментов Python.

**Контрольные вопросы:**

8. Постановка задачи классификации данных. Что такое бинарная классификация?
9. Общий алгоритм решения задачи классификации данных.
10. Чем отличаются обучающая и тестовая выборки? Какие существуют способы формирования обучающей и тестовой выборок?
11. Как рассчитываются значения полноты и точности?
12. Как рассчитывается значение площади под кривой ошибок и что оно характеризует?
13. Что показывает и как рассчитывается матрица ошибок?
14. Алгоритм и особенности метода k-ближайших соседей.
15. Алгоритм и особенности метода случайного леса.

#### **Лабораторная работа №4. Методы линейной регрессии.**

**Цель работы:** понять и научиться применять метод линейной регрессии в машинном обучении.

##### **Контрольные вопросы:**

1. Какие показатели применяются для измерения связи между признаками?
2. Как вычисляют коэффициент фенотипической корреляции в малых выборках?
3. Как вычисляют коэффициент фенотипической корреляции в больших выборках?
4. В чем заключается различие связи между признаками при положительных и отрицательных значениях коэффициента корреляции?
5. Как вычисляется коэффициент корреляции для альтернативных признаков?
6. В каких случаях используется коэффициент ранговой корреляции?
7. Как вычисляют коэффициент генетической корреляции?
8. Что характеризуют коэффициенты регрессии?

#### **Раздел 4. Логические методы.**

##### **Лабораторная работа №5. Базовые методы интеллектуального анализа данных.**

**Цель работы:** ознакомиться с возможностями классификации данных с помощью аналитического пакета Deductor Academic.

##### **Контрольные вопросы:**

1. Для чего служат алгоритмы g-mean и k-mean?
2. Какие алгоритмы обучения нейронной сети предлагает программный комплекс Deductor Academic?
3. В чем их отличие?
4. Что такое обучение с учителем? 5. Что такое обучение без учителя?

**Лабораторная работа №6. Методы машинного обучения, построенном на деревьях решений.**

**Цель работы:** познакомить обучающихся с методом машинного обучения, построенном на деревьях решений, а также научить строить сами деревья.

##### **Контрольные вопросы:**

1. Что представляет собой дерево решений?
2. Из каких объектов состоит дерево решений?
3. В чем отличие узла от листа?
4. Для каких задач Data Mining может использоваться дерево решений?
5. Какой вид правил используется в деревьях решений?
6. Всегда ли дерево, распознавшее все обучающие примеры, является наилучшим?
7. Какие существуют способы упрощения деревьев решений?

8. Почему узлы и листья, содержащие всего несколько примеров, имеет смысл отсекаать?
9. Дерево решений как линейный классификатор.
10. Что такое таблица сопряженности и как ее использовать для определения надежности классификации деревом решений.

## **Раздел 5. Особенности реальных данных.**

**Лабораторная работа №7.** Анализ признаков и оценка их информативности.

**Цель работы:** ознакомиться с возможностями аналитического пакета Deductor Academic.

### **Контрольные вопросы:**

1. Для чего служит программа Deductor Academic?
2. Зачем нужна предобработка данных?
3. Что такое парциальная предобработка?
4. Что такое вейвлет?
5. Какие данные можно импортировать в программу?

## **Раздел 6. Кластеризация данных.**

**Лабораторная работа №8.** Оптимизация и переобучение на примере метода "К-соседей".

**Цель работы:** исследование методов оптимизации и переобучения на примере метода "К-соседей".

### **Контрольные вопросы:**

11. Какие факторы влияют на переобучение?
12. Какие есть способы оценки переобучения?
13. Какие есть способы борьбы с переобучением?

**Лабораторная работа №9.** Изучение опыта применения методов кластеризации данных.

**Цель работы:** изучение современных приложений методов кластеризации данных в области прикладной информатики и программной инженерии на примере зарубежного опыта и зарубежных публикаций.

### **Контрольные вопросы:**

1. В чем состоит цель кластеризации? Приведите формальную постановку задачи кластеризации.
2. Какую проблему и из какой области решают авторы в статье? Зачем нужно решить эту проблему?
3. Как решалась эта проблема раньше: на основе информации из статьи (должна быть таблица методов решения с указанием источников в квадратных скобках и недостатков)?
4. Что предложили авторы нового в решении поставленной проблемы и для устранения какого недостатка (один абзац)?
5. С помощью каких известных методов кластеризации решается поставленная проблема в статье? Какова схема проверки качества решения?
6. Какие данные были использованы для проведения экспериментов?
7. Источники данных, количество и характеристики, примеры.
8. Каков итог решения проблемы и какие задачи требуется решать в будущем?
9. Какие недостатки приведенного в статье исследования Вы заметили?

**Лабораторная работа №10.** Программирование методов кластеризации данных.

**Цель работы:** изучить методы кластеризации, предложить модификацию выбранного метода и получить навыки в создании приложений для решения практической задачи анализа данных на основе методов кластеризации.

**Контрольные вопросы:**

10. В чем состоит цель кластеризации? Приведите формальную постановку задачи кластеризации.

11. Приведите перечень и особенности методов кластеризации.

12. Какие метрики применяют в кластеризации?

13. Приведите математические характеристики кластеров и меры

14. качества результатов кластеризации.

15. Охарактеризуйте этапы кластерного анализа выбранного метода.

16. Чем Ваше решение отличается от стандартного кластерного анализа на основе выбранного метода?

17. Сформулируйте выводы и объясните полученные результаты лабораторной работы.

**Лабораторная работа №11.** Лингвистическое резюмирование результатов кластеризации данных.

**Цель работы:** изучение и получение навыков в создании лингвистического описания (резюмирования) результатов кластеризации.

**Контрольные вопросы:**

1. В чем состоит цель лингвистического описания данных? Приведите

2. формальную постановку этой задачи.

3. Приведите определение лингвистической шкалы.

4. Приведите этапы создания лингвистической шкалы.

5. Сформулируйте выводы и объясните полученные результаты

6. лабораторной работы.

**Раздел 7. Нейронные сети.**

**Лабораторная работа №12.** Классификация с помощью перцептрона.

**Цель работы:** изучение модели нейрона перцептрона и архитектуры перцептронной однослойной нейронной сети.

**Контрольные вопросы:**

1. Структура перцептронного нейрона.

2. Правило нахождения количества нейронов в перцептронной сети для распознавания заданного числа классов.

3. Процесс обучения перцептрона.

**Критерии оценки**

**4 – 5 баллов:**

- выполнены все задания практической (лабораторной) работы,
- обучающийся четко и без ошибок ответил на все контрольные вопросы.

**3-4 баллов:**

- выполнены все задания практической (лабораторной) работы;
- обучающийся ответил на все контрольные вопросы с замечаниями.

**2-3 баллов:**

- выполнены все задания практической (лабораторной) работы с замечаниями;
- обучающийся ответил на все контрольные вопросы с замечаниями.

**0-2 баллов:**

- обучающийся не выполнил или выполнил неправильно задания практической (лабораторной) работы;
- обучающийся ответил на контрольные вопросы с ошибками или не ответил на контрольные вопросы.

**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ  
ФЕДЕРАЦИИ**

Федеральное государственное бюджетное  
образовательное учреждение высшего образования  
«ТЮМЕНСКИЙ ИНДУСТРИАЛЬНЫЙ УНИВЕРСИТЕТ»

**Задания в тестовой форме  
по дисциплине «Модели и методы интеллектуального анализа»**

**Тестовые материалы представлены в системе Eduson.**

**Раздел 1. Введение, основные понятия анализа данных.**

1. Выберите верные утверждения

- a) Одна из задач машинного обучения – научиться делать прогнозы для признаков
- b) Объекты описываются с помощью признаков
- c) Одна из задач машинного обучения – научиться делать прогнозы для объектов
- d) Признаки описываются с помощью объектов

2. Какие из этих задач являются задачами классификации?

- a) Прогноз температуры на следующий день
- b) Разделение книг, хранящихся в электронной библиотеке, на научные и художественные
- c) Поиск групп похожих пользователей интернет-магазина
- d) Прогноз оценки студента по пятибалльной шкале на экзамене по машинному обучению в следующей сессии

3. Какие свойства данных препятствуют однозначному построению разделяющей поверхности?

- a) Ортогональность
- b) Мультиколлинеарность
- c) Противоречивость
- d) Категориальность

4. Какая способность людей и систем позволяет получать им новые знания по наблюдению отдельных прецедентов (примеров)?

- a) Корректировать ошибку
- b) Обобщать
- c) Запоминать
- d) Распознавать образы

5. Какая задача лучше всего подходит под следующее описание. Нахождение такой функции  $F$ , которая бы наилучшим образом отображала неизвестные ранее объекты  $X$  в конечное множество целочисленных номеров (имен, меток), на основании обучающих пар  $(X, Y)$ ?

- a) Прогнозирование денежных затрат
- b) Кластеризация клиентов
- c) Классификация образов
- d) Выявление особенностей в данных

6. Почему для обучения моделей используются такие методы, как Градиентный спуск?

- a) Потому что метод позволяет корректировать параметры модели постепенно

- b) Потому что аналитические решения не всегда дают корректное решение
- c) Потому что такой подход позволяет получать более точные решения (Глобальный экстремум в отличие от локального)
- d) Потому что при большой размерности входных данных подобные методы работают быстрее

6. Выберите верные утверждения

- a) Метод Байеса – это во многом классический подход к классификации, основанный на оценке частоты встреч объектов со схожими признаками
- b) Благодаря универсальности статистического подхода метод Байеса позволяет решать любые задачи без априорной информации
- c) Данный метод позволяет очень хорошо обобщать высокоуровневые признаки
- d) Закон, задающий распределение вероятностей, который используется в предсказательной модели, сильно влияет на способ обобщения

### **Раздел 3. Линейная регрессия и классификация.**

1. Что такое линейная регрессия?

- 1) это функциональная зависимость, которая позволяет прогнозировать изменения непрерывных числовых параметров;
- 2) модель зависимости непрерывной переменной  $y$  от объясняющих ее факторов, в которой функция зависимости является линейной
- 3) модель зависимости дискретной переменной  $y$  от объясняющих ее факторов, в которой функция зависимости является линейной

2. Основная характеристика задачи бинарной классификации:

- 1 классификация осуществляется по одному признаку
- 2 зависимая переменная может принимать только два значения
- 3 классификация осуществляется по двум признакам

3. Классификация относится к стратегии:

- 1) обучения без учителя
- 2) обучения с учителем
- 3) оба ответа неверны

4. Явление переобучения характеризуется ...

- 1) чрезмерно точным соответствием модели конкретному набору обучающих примеров, при котором модель теряет способность к обобщению
- 2) возникновением, в случае слишком долгого обучения, недостаточного числа обучающих примеров или слишком сложной структуры модели
- 3) возникновением, в случае слишком долгого обучения, слишком сложной структуры модели

5. Задача классификации сводится к ...

- 1) нахождению частых зависимостей между объектами или событиями;
- 2) определению класса объекта по его характеристикам;
- 3) определению по известным характеристикам объекта значение некоторого его параметра;
- 4) поиску независимых групп и их характеристик в всем множестве анализируемых данных.

6. Множество примеров, используемое для конструирования модели, называется...

- 1) обучающим множеством

- 2) тестовым множеством
- 3) валидационным множеством

7. Коэффициент корреляции указывает:

- 1) на наличие связи
- 2) на отсутствие связи
- 3) на наличие связи и находится в интервале  $[-1;1]$
- 4) равен 0, если существует связь между изучаемыми явлениями
- 5) нет правильного ответа

8. Проверить значимость параметров уравнения регрессии можно, используя:

- 1) t-статистику
- 2) коэффициент детерминации
- 3) коэффициент корреляции
- 4) коэффициент ковариации
- 5) нет правильного ответа

1. Коэффициент уравнения регрессии показывает

- 1) На сколько % изменится результат при изменении фактора на 1 %.
- 2) На сколько % изменится фактор при изменении результата на 1 %.
- 3) На сколько ед. изменится результат при изменении фактора на 1 ед.
- 4) На сколько ед. изменится фактор при изменении результата на 1 ед.
- 5) Во сколько раз изменится результат при изменении фактора на 1 ед.

1. Выбор вида функциональной зависимости в уравнении регрессии называется:

- 1) агрегированием модели
- 2) параметризацией модели
- 3) линеаризацией модели
- 4) структуризацией модели
- 5) спецификацией модели

1. Коэффициент эластичности показывает

- 1) На сколько ед. изменится фактор при изменении результата на 1 ед.
- 2) На сколько ед. изменится результат при изменении фактора на 1 ед.
- 3) Во сколько раз изменится результат при изменении фактора на 1 ед.
- 4) На сколько % изменится результат при изменении фактора на 1 %.
- 5) На сколько % изменится фактор при изменении результата на 1 %.

1. Если эконометрическая модель содержит только одну объясняющую переменную, то она имеет название:

- 1) парной линейной регрессии;
- 2) парной регрессии;
- 3) парной нелинейной регрессии;
- 4) множественной линейной регрессии;
- 5) множественной регрессии.

### **Раздел 6. Кластеризация данных.**

1. Задача кластеризации заключается в ...

- 1) нахождения частых зависимостей между объектами или событиями;
- 2) определения класса объекта по его характеристиками;
- 3) определение по известным характеристиками объекта значение некоторого его параметра;

- 4) поиска независимых групп и их характеристик в всем множестве анализируемых данных.
  
2. Целью поиска ассоциативных правил является ...
  - 1) нахождения частых зависимостей между объектами или событиями;
  - 2) определения класса объекта по его характеристиками;
  - 3) определение по известным характеристиками объекта значение некоторого его параметра;
  
3. До предполагаемых моделей относятся такие модели данных:
  - 1) модели классификации и последовательностей;
  - 2) регрессивные, кластеризации, исключений, итоговые и ассоциации;
  - 3) классификации, кластеризации, исключений, итоговые и ассоциации;
  - 4) модели классификации, последовательностей и исключений.
  
4. В описательных моделей относятся следующие модели данных:
  - 1) модели классификации и последовательностей;
  - 2) регрессивные, кластеризации, исключений, итоговые и ассоциации;
  - 3) классификации, кластеризации, исключений, итоговые и ассоциации;
  - 4) модели классификации, последовательностей и исключений.
  
5. Модели классификации описывают ...
  - 1) правила или набор правил в соответствии с которыми можно отнести описание любого нового объекта к одному из классов;
  - 2) функции, которые позволяют прогнозировать изменения непрерывных числовых параметров;
  - 3) функциональные зависимости между зависимыми и независимыми показателями и переменными в понятной человеку форме;
  - 4) группы, на которые можно разделить объекты, данные о которых подвергаются анализа.
  
6. Модели последовательностей описывают ...
  - 1) правила или набор правил в соответствии с которыми можно
  - 2) отнести описание любого нового объекта к одному из классов;
  - 3) функции, которые позволяют прогнозировать изменения непрерывных
  - 4) числовых параметров;
  - 5) функциональные зависимости между зависимыми и независимыми показателями и переменными в понятной человеку форме;
  - 6) группы, на которые можно разделить объекты, данные о которых подвергаются анализа.
  
7. Регрессивные модели описывают ...
  - 1) правила или набор правил в соответствии с которыми можно отнести описание любого нового объекта к одному из классов;
  - 2) функции, которые позволяют прогнозировать изменения непрерывных
  - 3) числовых параметров;
  - 4) функциональные зависимости между зависимыми и независимыми показателями и переменными в понятной человеку форме;
  - 5) группы, на которые можно разделить объекты, данные о которых подвергаются анализа.
  
8. Кластер можно охарактеризовать как ...

- 1) группу объектов, имеющих общие свойства
  - 2) один объект, изолированный от других
  - 3) группу объектов, имеющую внутреннюю однородность
9. Метод деревьев решений применяется для решения задач ...
- 1) классификации
  - 2) кластеризации
  - 3) классификации и кластеризации
10. Иерархические алгоритмы применяются для решения задач ...
- 1) классификации
  - 2) кластеризации
  - 3) классификации и кластеризации

### **Раздел 7. Нейронные сети.**

1. Выберите верные утверждения:
  - 1) ИНС проще подобрать под любую нелинейную задачу. Все, что нужно сделать, это увеличивать число слоев пропорционально числу признаков
  - 2) ИНС позволяют обрабатывать более высокоуровневые признаки за счет нелинейной функции активации и последовательным слоям
  - 3) По сравнению с Регрессией ИНС практически не подвержены Переобучению при любом количестве нейронов
  - 4) С точки зрения математического аппарата ИНС – это комбинация полиномиальной регрессии высокого порядка и формулы Байеса
  - 5) ИНС может аппроксимировать любую нелинейную непрерывную функцию, но это еще не гарантирует 100% сходимости на произвольных данных
  - 6) ИНС в отличие от регрессии может хорошо обрабатывать высокую степень мультиколлинеарности и противоречивости в данных
2. Сеть какого типа лучше использовать для прогнозирования временных рядов?
  - 1) Сверточную
  - 2) ART MAP
  - 3) Импульсную
  - 4) MLP
  - 5) Рекуррентную
  - 6) Когнитрон
3. Сеть какого типа лучше использовать для обработки трехмерных сцен?
  - 1) MLP
  - 2) Рекуррентную
  - 3) ART MAP
  - 4) Сверточную
  - 5) Когнитрон
  - 6) Импульсную
4. Сеть какого типа лучше использовать для решения задачи классификации клиентов по одиночному вектору клиентских характеристик (с учетом того, что этот вектор содержит большое количество категориальных признаков)?
  - 1) Автокодировщик
  - 2) MLP
  - 3) Когнитрон

- 4) ART MAP
  - 5) Сверточную
  - 6) Рекуррентную
  - 7) Импульсную
5. Сеть какого типа можно использовать в условиях постоянного изменения данных, когда точной выборки еще не существует и сеть приходится постоянно дообучивать на новых классах?
- 1) MLP
  - 2) Сверточную
  - 3) Когнитрон
  - 4) Рекуррентную
  - 5) ART MAP
  - 6) Автокодировщик
  - 7) Импульсную
6. Что означает величина OUT?
- 1) выход суммирующего блока
  - 2) значение активационной функции
  - 3) входной сигнал нейрона
  - 4) выходной сигнал нейрона
7. Активационной функцией называется:
- 1) функция, вычисляющая выходной сигнал нейрона
  - 2) функция, суммирующая входные сигналы нейрона
  - 3) функция, корректирующая весовые значения
  - 4) функция, распределяющая входные сигналы по нейронам
8. Матричное умножение  $XW$  вычисляет:
- 1) выходной нейронный сигнал
  - 2) выход суммирующего блока
  - 3) входной нейронный сигнал
  - 4) вход суммирующего блока
9. Персептроном Розенблатта называется:
- 1) однослойная нейронная сеть с бинарным входом
  - 2) однейронная сеть
  - 3) многослойная сеть прямого распространения
10. Сколько слоев имеет персептрон Розенблатта?
- 1) один
  - 2) два
  - 3) три
  - 4) любое конечное число
11. Выходом персептрона являются:
- 1) значения отрезка  $[0;1]$
  - 2) значения отрезка  $[-1;1]$
  - 3) вся числовая ось  $(-\infty;+\infty)$
12. Входом персептрона являются:
- 1) значения 0 и 1

- 2) вся действительная ось  $(-\infty; +\infty)$
  - 3) вектор, состоящий из нулей и единиц
  - 4) вектор, состоящий из действительных чисел
13. Какая активационная функция используется в персептроне?
- 1) пороговая
  - 2) гиперболического тангенса
  - 3) сигмоидальная
14. Способность персептрона решать определенную задачу называется:
- 1) способностью к обучению
  - 2) представимостью
  - 3) линейной разделимостью
15. Персептронной представимостью называется:
- 1) способность персептрона моделировать определенную функцию
  - 2) способность персептрона к обучению
  - 3) разделение гиперпространства решений гиперплоскостью
16. Однонейронный персептрон с двумя входами:
- 1) разделяет плоскость  $XOY$  на две полуплоскости
  - 2) разделяет трехмерное пространство  $XOY$  на два полупространства
  - 3) выделяет замкнутую область
17. Обучением называют:
- 1) процедуру подстройки весовых значений
  - 2) процедуру подстройки сигналов нейронов
  - 3) процедуру вычисления пороговых значений для функций активации
18. Что называется обучающей выборкой для обучения персептрона?
- 1) набор входных векторов, для которых заранее известны значения аппроксимируемой функции
  - 2) набор выходных векторов, являющихся точными значениями аппроксимируемой функции
  - 3) набор пар входов и выходов, используемых при обучении
19. Нейронная сеть является обученной, если:
- 1) алгоритм обучения завершил свою работу и не заиклился
  - 2) при запуске обучающих входов она выдает соответствующие обучающие выходы
  - 3) при подаче на вход некоторого вектора сеть будет выдавать ответ, к какому классу векторов он принадлежит
20. Запускаем обучающий вектор  $X$ . В каком случае весовые значения не нужно изменять?
- 1) если на выходе сеть даст 0
  - 2) если на выходе сеть даст 1
  - 3) если сигнал персептрона совпадает с правильным ответом
21. Теорема о сходных персептронах утверждает, что:
- 1) если данная задача представляет персептрон, то он способен ей обучиться
  - 2) алгоритм обучения всегда сходится
  - 3) найдутся задачи, которым персептроны не смогут обучиться

22. Теорема о "зацикливании" персептрона утверждает, что:
- 1) любой алгоритм обучения зацикливается
  - 2) если данная задача не представима персептроном, то алгоритм обучения зацикливается
  - 3) если задача не имеет решения, то алгоритм обучения зацикливается
23. В каком случае персептрон может обучиться решать данную задачу?
- 1) если задача представима персептроном
  - 2) если задача имеет решение
  - 3) если задача имеет целое численное решение
24. Когда алгоритм обучения персептрона зацикливается?
- 1) если данная задача непредставима персептроном
  - 2) если данная задача не имеет решения
  - 3) если коэффициенты в алгоритме обучения подобраны неверно
25. Можем ли мы за конечное число шагов после запуска алгоритма обучения персептрона сказать, что персептрон не может обучиться данной задаче?
- 1) да
  - 2) нет
  - 3) в зависимости от задачи
26. Если данный персептрон заменить персептроном с целочисленными весами, то:
- 1) новый персептрон будет решать более узкую задачу
  - 2) новый персептрон будет решать ту же самую задачу
  - 3) новый персептрон будет решать более широкую задачу
27. Вопрос о выборе шага при применении процедуры обучения решается следующим образом:
- 1) веса и порог следует изменять на 1
  - 2) веса и порог следует изменять на число  $\leq 1$
  - 3) веса и порог следует изменять на целое число
28. Теорема о двухслойности персептрона утверждает, что:
- 1) любой многослойный персептрон может быть представлен в виде двухслойного персептрона
  - 2) в любом многослойном персептроне могут обучаться только два слоя
  - 3) способностью к обучению обладают персептроны, имеющие не более двух слоев

### **Критерии оценки за аттестацию**

**0-2 баллов** - менее 60% правильных ответов.

**3 балла** - от 61% до 75% правильных ответов.

**4 балла** - от 76% до 91% правильных ответов.

**5 баллов** - 100% правильных ответов.

**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ  
ФЕДЕРАЦИИ**

Федеральное государственное бюджетное  
образовательное учреждение высшего образования  
«ТЮМЕНСКИЙ ИНДУСТРИАЛЬНЫЙ УНИВЕРСИТЕТ»

**Комплект заданий для домашней работы  
по дисциплине «Модели и методы интеллектуального анализа»****Раздел 3. Линейная регрессия и классификация.  
Домашняя работа №1**

**Задание:** Ответьте на вопросы о данных по авиарейсам в США за январь-апрель 2008 года. По ссылке расположены

Данные и их описание

1) Считайте выборку из файла при помощи функции `pd.read_csv` и ответьте на следующие вопросы:

- Имеются ли в данных пропущенные значения?
- Сколько всего пропущенных элементов в таблице "объект-признак"?
- Сколько объектов имеют хотя бы один пропуск?
- Сколько признаков имеют хотя бы одно пропущенное значение?

2) Преобразуйте каждый признак `FeatureName` из указанных в пару новых признаков `FeatureName_Hour`, `FeatureName_Minute`, разделив каждое из значений на часы и минуты. Не забудьте при этом исключить исходный признак из выборки. В случае, если значение признака отсутствует, значения двух новых признаков, его заменяющих, также должны отсутствовать.

3) Некоторые из признаков, отличных от целевой переменной, могут оказывать чересчур значимое влияние на прогноз, поскольку по своему смыслу содержат большую долю информации о значении целевой переменной. Изучите описание датасета и исключите признаки, сильно коррелирующие с ответами. Ваш выбор признаков для исключения из выборки обоснуйте.

4) Приведите данные к виду, пригодному для обучения линейных моделей. Для этого вещественные признаки надо отмасштабировать, а категориальные — привести к числовому виду. Также надо устранить пропуски в данных. Реализуйте функцию `transform_data`, которая принимает на вход `DataFrame` с признаками и выполняет следующие шаги: – Замена пропущенных значений на нули для вещественных признаков и на строки 'nan' для категориальных. – Масштабирование вещественных признаков с помощью `StandardScaler`. – One-hot-кодирование категориальных признаков с помощью `DictVectorizer` или функции `pd.get_dummies`. Метод должен возвращать преобразованный `DataFrame`, который должна состоять из масштабированных вещественных признаков и закодированных категориальных (исходные признаки должны быть исключены из выборки).

5) Разбейте выборку и вектор целевой переменной на обучение и контроль в отношении 70/30 (для этого можно использовать функцию `train_test_split`).

6) Обучите линейную регрессию на 1000 объектах из обучающей выборки и выведите значения `MSE` и `R2` и на этой подвыборке и контрольной выборке. Проинтерпретируйте полученный результат — насколько качественные прогнозы строит полученная модель? Какие проблемы наблюдаются в модели?

7) Обучите линейные регрессии с L1- и L2-регуляризатором, подобрав лучшее значение параметра регуляризации при помощи кросс-валидации. Выведите значения и на обучающей и контрольной выборках. Удалось ли решить указанные вами ранее проблемы?

**Раздел 6. Кластеризация данных.**

## Домашняя работа №2

### Применение алгоритма кластеризации: самоорганизующиеся карты Кохонена

#### Задание:

1. Импортируйте в АП «Deductor» исходные данные из файла C:\Program\Files\BaseGroup\Deductor\Samples\CreditSample.txt.
2. Процесс построения карты Кохонена состоит из 10 этапов.
3. Запустите мастер обработки, в котором в разделе «Data Mining» выберете способ обработки данных «Карта Кохонена».
4. Запустите процесс построения карты Кохонена.
5. Проведите настройку отображения карты.
6. Проанализируйте результаты, что можно сказать о вероятности возврата кредита для групп 2, 3 и 4?
7. Используя различные отображения карты Кохонена, постройте 3-4 правила выдачи кредитов.
8. Ответьте на вопросы: – для чего используются карты Кохонена? – по какому принципу происходит перенос многомерного пространства на пространство меньшей размерности?

#### Критерии оценки

##### Критерии оценки домашнего задания:

##### Баллы

- соответствие предполагаемым ответам;	
- правильное использование алгоритма выполнения действий (методики, технологии и т.д.);	0-3
выставляется студенту, если он владеет категориальным аппаратом, может привести формулы расчета, рассчитать задание;	4-5
выставляется студенту, если он владеет категориальным аппаратом, может привести классификацию факторов явления, решить поставленную задачу и проанализировать полученные результаты;	6-7
выставляется студенту, если он владеет категориальным аппаратом, может привести классификацию факторов явления, решить поставленную задачу и проанализировать полученные результаты, объяснить причины отклонений от желаемого результата, отстоять свою точку зрения, приводя факты;	8-10
<b>Максимальное количество баллов</b>	<b>10</b>

**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ  
ФЕДЕРАЦИИ**

Федеральное государственное бюджетное  
образовательное учреждение высшего образования  
«ТЮМЕНСКИЙ ИНДУСТРИАЛЬНЫЙ УНИВЕРСИТЕТ»

**Задания для коллоквиума  
по дисциплине «Модели и методы интеллектуального анализа»****Раздел 5. Особенности реальных данных.**

1. Понятие метаданных и их роль в системе поддержки принятия решений. Разновидности метаданных по функциональным областям и их использование в деловой деятельности.
2. Общие методы структурирования информации в процессе принятия решений. Информационное моделирование корпоративных данных. Бизнес-размерности и бизнес-метрики.
3. Специфические характеристики аналитических моделей данных. Взаимосвязь с реляционной моделью данных. Таблицы фактов и таблицы размерностей. Процесс перехода от информационных требований к разработке модели данных.
4. Классификация данных как аспект информационной безопасности
5. Методы многомерного анализа данных: методы многомерного корреляционного анализа
6. Основные отличия технологии OLAP (On Line Analytical Processing) от традиционных способов анализа данных. Инструменты и методы аналитической обработки данных: гиперкуб, свертка, поворот, конкретизация, получение слоя.
7. Различные модели аналитической обработки данных: MOLAP, ROLAP, HOLAP. Методы реализации OLAP-систем.
8. Основные понятия технологии DataMining. Специфика анализируемых данных: разреженные, отсутствующие и неточные данные. Разведочный анализ данных, анализ тенденций. Предсказательные и описательные модели.
9. Преимущества использования технологии DataMining при выработке стратегических решений и примеры из реальных прикладных областей.
10. Методы использования мультимедийной информации в технологиях Хранилищ данных и добычи данных.

**Критерии оценки**

**9 – 10 баллов** - выставляется, если обучающийся раскрыл содержание материала в объеме, предусмотренном программой, изложил материал грамотным языком в определенной логической последовательности, точно используя терминологию данного предмета как учебной дисциплины; отвечал самостоятельно без наводящих вопросов преподавателя. Возможны одна – две неточности при освещении второстепенных вопросов или в выкладках, которые обучающийся легко исправил по замечанию преподавателя.

**7-8 баллов** - выставляется, если ответ обучающегося удовлетворяет в основном требованиям на отметку «10-8» баллов, но при этом имеет место один из недостатков: допущены одна - две неточности при освещении основного содержания ответа, исправленные по замечанию преподавателя; допущены ошибка или более двух неточностей при освещении второстепенных вопросов или в выкладках, легко исправленные по замечанию преподавателя.

**4-6 баллов** - выставляется в следующих случаях: неполно или непоследовательно раскрыто содержание материала, имеются ошибки при ответах на вопросы, , но показано общее

понимание вопроса и продемонстрированы умения, достаточные для дальнейшего усвоения материала, определенного учебной программой дисциплины.

**0-3 баллов** - выставляется в случаях, если не раскрыто основное содержание учебного материала; обнаружено незнание или неполное понимание обучающимся большей или наиболее важной части учебного материала; допущены грубые ошибки при ответах на вопросы опроса, допущены ошибки в определении понятий при использовании специальной терминологии в рисунках, схемах, выкладках, которые не исправлены после нескольких наводящих вопросов преподавателя.

**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ  
ФЕДЕРАЦИИ**

Федеральное государственное бюджетное  
образовательное учреждение высшего образования  
«ТЮМЕНСКИЙ ИНДУСТРИАЛЬНЫЙ УНИВЕРСИТЕТ»

**Вопросы к экзамену**

по дисциплине «Модели и методы интеллектуального анализа»

11. Линейная регрессионная модель. Метод наименьших квадратов. Основные свойства оценок метода наименьших квадратов.
12. Проверка линейных гипотез в рамках классической модели регрессии.
13. Дискриминантный анализ: постановка задачи и ее решение в случае известных параметров. 4. Кластерный анализ: постановка задачи и основные понятия.
14. Спектральный анализ временных рядов. Периодограмма выборочная. Свойства периодограммы. Сглаженная периодограмма.
15. Линейная ARMA(p,q) модель временных рядов. Стационарность, обратимость. Автокорреляционная и частная автокорреляционная функции. Оценивание параметров. Прогнозирование.
16. Общие принципы выбора страховых тарифов и тарифных ставок.
17. Асимптотическая формула для страховых тарифов в статической модели страхования: сравнение классической и факторизационной модели.
18. Гарантированные оценки страховых тарифов для факторизационной модели страхования при пуассоновском объеме портфеля.
19. Преобразование Фурье и его свойства. Линейные фильтры. Теорема о свертке.
20. Кратномасштабный анализ. Алгоритмы разложения и реконструкции.
21. Пороговая обработка коэффициентов вейвлет-разложения. Методы выбора порога.
22. Модель вычислений MapReduce. Пример реализации (Google MapReduce либо Hadoop MapReduce). Стадии вычислений, основные компоненты, понятие локальности данных.
23. Модель обработки данных в Apache Spark. Понятие resilient distributed dataset (RDD), примеры преобразований и действий над RDD. Граф вычислений.
24. Распределенная файловая система. Пример распределенной файловой системы (Google File System либо HDFS). Компоненты, основные операции, сценарии чтения и записи данных, механизмы отказоустойчивости.
25. Способы объектно-реляционного отображения для классов и атрибутов, бинарных и N-арных ассоциаций, классов ассоциаций, иерархий наследования. Примеры применения этих способов. Моделирование схемы реляционной базы данных с помощью диаграммы классов.
26. Образцы (паттерны) проектирования, их классификация и способ описания. Примеры образцов: структурного, поведенческого и порождающего.
27. Модели прогнозирования на основе деревьев решений. Алгоритмы CHAID, CART, C4.5: критерии поиска разбиений, параметры ограничения роста и обрубания дерева.
28. Нейронные сети прямого распространения. Архитектуры MLP и RBF: структура сетей, виды функций активации, алгоритмы обучения, борьба с переобучением и с проблемой локальных минимумов.
29. Поиск ассоциативных правил. Алгоритмы Apriori и FP-tP-tree: построение частых эпизодов с ограничением по поддержке и формирование правил с ограничением по достоверности.
30. rename, end). Использование by-группировки и retain переменных для расчета агрегаций.
31. Преобразование изображения на основе эквализации гистограммы яркости.

32. Свёртка функций. Одномерная и двумерная свёртка и её свойства. Дискретная свёртка изображений. Обработка края изображения при свёртке.
33. Последовательная и параллельная сложность алгоритмов, информационный граф и ресурс параллелизма алгоритмов.
34. Архитектурные особенности графических процессоров, направленные на массивно-параллельные вычисления.

### **Критерии оценки**

**100 – 91 баллов:** студент обладает системными теоретическими знаниями (знает методику выполнения практических навыков, показания и противопоказания, возможные осложнения, нормативы и проч.), без ошибок самостоятельно демонстрирует выполнение практических умений;

90-76 баллов: студент обладает теоретическими знаниями (знает методику выполнения практических навыков, показания и противопоказания, возможные осложнения, нормативы и проч.), самостоятельно демонстрирует выполнение практических умений, допуская некоторые неточности (малозначительные ошибки), которые самостоятельно обнаруживает и быстро исправляет;

**75-61 баллов:** студент обладает удовлетворительными теоретическими знаниями (знает основные положения методики выполнения практических навыков, показания и противопоказания, возможные осложнения, нормативы и проч.), демонстрирует выполнение практических умений, допуская некоторые ошибки, которые может исправить при коррекции их преподавателем;

**60-0 баллов:** студент не обладает достаточным уровнем теоретических знаний (не знает методики выполнения практических навыков, показаний и противопоказаний, возможных осложнений, нормативы и проч.) и/или не может самостоятельно продемонстрировать практические умения или выполняет их, допуская грубые ошибки.